Appendix A
Review of the Bell Atlantic North and South CPR Statistical Audit Plan

This appendix provides the supporting details underlying Ernst & Young's findings in its analysis of the sampling plans used by the FCC for each of the Current Property Records (CPR) audits of Bell Atlantic North (the former NYNEX telephone companies) and South (the original Bell Atlantic telephone companies). It should be noted that the FCC has modified some of the values in the tables since their initial draft report. However, the scoring changes do not affect any of the discussion or conclusions of this appendix, which still reflects the originally reported numbers.

In what follows, we give a more in-depth discussion of issues that are raised in our summary report on the audit. Specifically, these issues are:

- a possibly inappropriate choice of sample design;
- the incorrect calculation of margins of error; and
- many sources of bias that affect the estimates.

A fourth issue, the lack of a two-way audit, was also discussed in the summary report. While we could go into the mathematics of how to produce an estimate of missing property from the results of a two-way audit, we do not feel that it is warranted here. Pointing out the failure to even attempt the necessary two-way audit should be enough at this point.

After discussing each of the above issues, we will provide a comparison of calculations we have made with those published in the FCC's draft report. To set the tone for these discussions, we first provide some definitions, and describe the notation that will be used in the equations that will follow.

In general, when Bell Atlantic scores are used, the one-sided lower 95 and 99 percent bounds for the in-place cost of non-locatable items are either below zero, or are a minute fraction of the Bell Atlantic company's total investment.

Definitions and Notation

Unless otherwise specified, all definitions and notation will refer to a singular database. This means that only one group of companies' (e.g. Bell Atlantic North or Bell Atlantic South) audit is considered at a time, not both simultaneously.

- The population of interest is the central office hardwire records of the current property record (CPR) database of either Bell Atlantic North as of 3/31/97 or Bell Atlantic South as of 3/31/97.

Denote the total number of records in this population by M_0.

- We shall refer to a record in the CPR database as a line-item.

- A central office location is denoted by the first eight characters of the Common Language Location Indicator (CLLI). We will henceforth refer to each central office location as a CLLI.

- For the audit, CLLIs are divided into L groups or strata.

- A line-item belongs to one, and only one CLLI.

- For each $h = 1,\ldots,L$ of the CLLI strata, let

  N_h = the number of CLLIs in stratum h,

  n_h = the number of CLLIs selected for the audit in a sample from stratum h,

  M_h = the total number of line-items across all CLLI in stratum h,

  M'_h = the total number of line-items in all CLLIs selected for the audit within stratum h, and

  m_h = the total number of line-items selected for the audit in stratum h.

  Note that $$M_0 = \sum_{h=1}^{L} M_h .$$

- Within stratum h ($h = 1,\ldots,L$), let

  M_{hi} = the number of records in CLLI i ($i = 1,\ldots,N_h$) of stratum h, and

  36 = the number of FCC sampled line-items in selected CLLI i ($= 1,\ldots,n_h$) of stratum h.

  Note that $$M_h = \sum_{i=1}^{N_h} M_{hi} ,$$

  $$M'_h = \sum_{i=1}^{n_h} M_{hi} \text{, and}$$

  $$m_h = 36 \cdot n_h .$$

- Within CLLI i ($i = 1,\ldots,N_h$ when referring to the whole population, or $i = 1,\ldots,n_h$ when referring to the sample of locations for the audit) of stratum h ($h = 1,\ldots,L$), let

y_{hij} denote the observed value for line-item j ($j = 1,\ldots,M_{hi}$ for the population of line-items, but $j = 1,\ldots,36$ for line-items chosen for the audit) within CLLI i of stratum h. For example:

> if you are interested in the number of compliant line-items, then y_{hij} is either 0 or 1 when a line-items is either non-compliant or compliant; or
>
> if you are interested in the total in-place cost for line-items that can't be located, then y_{hij} is the in-place cost of a line-item that cannot be located, and zero otherwise.

Sample Design Considerations

A sample design is the plan for choosing items for a sample. According to the draft report, the CPR hardwire audit used a two-stage, stratified cluster design. This was accomplished via the following steps:

1. The total number of hardwire line-items for the audit sample was determined to be 1082.
   - The methodology for determining this assumed a simple random sample would be taken at both stages.
   - The criteria for determining the sample size was a desire to have a margin of error for the proportion of compliant line-items of at most 0.025.
   - It appears to have been implicitly assumed that the degrees of freedom of the estimator would be large enough to use normal distribution theory.
2. It was determined that auditors would try to find the property corresponding to 36 randomly chosen line-items within each of a randomly selected central office locations.
3. The number of central office locations needed for the audit was determined to be 30, the result of dividing 1082 by 36 and truncating to an integer.
4. The sampling frame was determine as follows.[1]
   - All line-items were clustered within locations that were determined by eight character CLLI codes.
   - After removing non-hardwire records, line-item counts were done for each CLLI.
   - CLLIs with less than a minimum number of line-items were discarded (79 was the minimum for the North, and 36 was the minimum for the South), and the remaining CLLIs were considered the central offices.
5. The CLLIs in the frame were divided into 11 and 12 strata, for the North and South respectively, based on the number of line-items.

---

[1] This is not described in the draft report. This procedure was described by the FCC staff to Bell Atlantic, and subsequently relayed to us.

6. The sample size of 30 CLLIs was allocated across the strata proportionately to the total number of records in each stratum.[2] After adjusting the resulting number to be integers that added up to thirty, any stratum that was allocated less than two CLLI selections had its allocation increased to two.[3] This increased the total number of CLLIs in the sample to 34 for the North and 32 for the South. In turn, this increased the total number of line-item for the audit to $34 \cdot 36 = 1{,}224$ (North) and $32 \cdot 36 = 1{,}152$ (South).
7. Within each stratum, CLLI were randomly selected according to the allocation plan in step 6.
8. For each CLLI selected in step 7, thirty-six line items were randomly selected for the audit.[4]

While this sample design can be used to calculate estimates of many different population quantities, most estimates produced from it will not have very good precision. Major decisions for the design were based on the desire for a precise estimate of the proportion of compliant records. These include:

- determining the total number of line-items for the audit;
- allocation of the total number of CLLIs across strata; and to some extent,
- the division of CLLIs into strata.

Even at that, the sample design does not produce the desired effect – a margin of error of at most 0.25 for the estimate the proportion of compliant line items. This is due to the fact that the effect of clustering – sampling line-items within a CLLI – was not taken into account.

Methods based on simple random sampling were used, but the design is more complex than a simple random sample. For an account of how to design a complex sample so that a planned precision can be approximately achieved see Chapter 8 of Kish.[5]

Furthermore, if a precise estimate of the total in-place cost associated with non-locatable line-items is desired, then the sample design should take this into account. Selecting CLLI proportional to the total in-place cost of each CLLI, and stratification based on in-

---

[2] The draft report states that Neyman allocation was used. It does not state what was used as each stratum's variance, S^2_h. We suspect that the variance of the proportion of all compliant line-items in the stratum was used with the proportion set at 0.5. If so, the variances are treated as being the same across all strata, and the allocation becomes proportionate to record counts. Our own calculations using proportionate record counts allocation produce results which are consistent with those published in the summary table on page 7 of the draft report's Appendix B.

[3] Based on a footnote in the draft report concerning advise from Census Bureau staff, and discussions with Bell Atlantic personal, we believe that the increase to 2 locations in a stratum was done after many of the first 30 locations had been visited, probably as a result of a review by Census staff.

[4] From footnote 18 of the draft reports Appendix B, we know that when the audit team arrived at the central office location, if it was determined that the property associated with a line-item was "too hard-to-get-to," another line item was substituted. This line item was the one that preceded the randomly selected item in the CPR listing. This has the potential to introduce bias into estimates.

[5] Kish, L. (1965). *Survey Sampling*. John Wiley & Sons, New York.

place cost are two concepts that may help reduce the variance of in-place cost related estimators. For more on audit sampling issues, see "Statistical Models and Analysis in Auditing."[6]

As a general rule, the precision of dollar value estimators is much more sensitive to design decisions than are proportion estimators. By this we mean that a design made for a precise dollar estimator will most likely produce a proportion estimate with acceptable precision. The reverse of this is seldom true. Additionally, more CLLIs need to be selected in order to use normal approximation theory. This issue will be discussed more fully in the next section.

Finally, there is mention in the report that after the sample of CLLI was chosen, a check was made to see that all the states that are served by a company had a CLLI in the sample. This suggests that the sample would not have been accepted if all states were not represented. If the FCC wanted to have all states represent, then they are using the wrong approach.

A second stratification should have been done on state, and the whole CLLI sample size allocated according to the two-way stratification. Just as with the single stratification, every stratum would need at least two CLLIs allocate too it. Since this was not done, the FCC will not be able to produce precise estimates at the state level, even though they made sure that every state was represented in the sample. The sample that the FCC has drawn is not representative by state.

Margin of error

The margin of error is a measure of the precision of an estimator. It is usually the plus/minus part of a confidence interval of the form:

$$\tilde{X} \pm t \cdot s(\tilde{X}),$$

where $\tilde{X}$ is an estimator of some population quantity X, e.g., the total number of compliant line-items in the CPR, or the total in-place cost associated with missing property. The quantity $s(\tilde{X})$ is the standard error of the estimator, and t is a multiplying factor that is determined by the distribution of the standardized quantity

$$\frac{\tilde{X} - X}{s(\tilde{X})},$$

and the confidence that one wants to have in the estimate. Typically, t is a percentile of the standard normal distribution or Student's t distribution. Most basic statistics books

[6] Panel on Nonstandard Mixtures of Distributions (1989). Statistical Models and Analysis in Auditing. *Statistical Science*, **4**, No. 1, pp. 2-33.

have tables for finding these values. Statistical software and spreadsheet programs can also be used.

Following the discussion in Cochran,[7] if $\tilde{X}$ has a normal distribution with mean X, and $s(\tilde{X})$ is well determined, then t comes from the standard normal distribution. These are two very important assumptions, and if they are not true, other types of error bounds need to be calculated using more advanced techniques.

The more well known situation occurs when $\tilde{X}$ has a normal distribution, but the sample size is not large enough for $s(\tilde{X})$ to be well determined. In this case, the degrees of freedom need to be considered, and Student's t distribution is used to find the multiplying factor.

In a situation where stratification has been used, one needs to consider the degrees of freedom provided by each stratum. The distribution of $s(\tilde{X})$ is in general too complicated to simply compute the degrees of freedom for each stratum in the usual way – taking the CLLI sample size within the stratum minus one, i.e., $(n_h - 1)$ – and then add them up across all strata. An approximate method of assigning an effective number of degrees of freedom to $s(\tilde{X})$ has been worked out by Satterthwaite.[8]

Let $v(\tilde{X})$ be the total variance of the estimator, and $v_h(\tilde{X})$ the component of $v(\tilde{X})$ from stratum h. Then the effective degrees of freedom is given by

$$n_e = \frac{\{v(\tilde{X})\}^2}{\sum_h \frac{\{v_h(\tilde{X})\}^2}{n_h - 1}} .$$

The value of n_e always lies between the smallest of the values $(n_h - 1)$ and their sum. For the audits described in the draft report, this value will lie between 1 and 23 for the North, and 1 and 20 for the South. These values are too small for the normal distribution to be used.

Why is it that the central limit theorem does not apply when there is a relatively large total sample size (1,224 or 1,152)? This is due to the two-stage design. The variance between CLLIs contributes much more towards total variance than the variances within CLLIs. Thus, the number of locations chosen plays an important role, and this number was chosen to be relatively small in the FCC sample design.

---

[7] See Cochran, W. G. (1967). *Sampling Techniques*, 3rd ed. Wiley, New York. Pp. 95 -96.
[8] Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics*, **2**, pp. 110-114.

In the calculation section below, we show that the effective degrees of freedom is in the range 1 to 14, depending on the estimation method used and the scoring of the property records audited.

The draft report uses the multiplying factor 1.96, obtained from the standard normal distribution for a 95 percent two-sided confidence level. The table below shows the multiplying factor associated with different confidence levels from Student's t distribution with different degrees of freedom.

Degrees of Freedom	**One Sided Confidence Bounds**			**Two Sided Confidence bounds**		
n_e	**90%**	**95%**	**99%**	**90%**	**95%**	**99%**
1	3.078	6.314	31.821	6.314	12.706	63.656
5	1.476	2.015	3.365	2.015	2.571	4.032
9	1.383	1.833	2.821	1.833	2.262	3.250
14	1.345	1.761	2.624	1.761	2.145	2.977

Notice that the multiplying factors for two-sided bounds at 95 percent confidence are larger than the value from the normal distribution – namely, 1.96. Thus, the reported margin of error for all estimates in the draft report needs to be increased.

The above analysis is only useful if the underlying distribution of the estimator is normally distributed. This is true probably for an estimator of the proportion of compliant records – although it should be confirmed. But it is not very likely that this will be true for estimators associated with dollar values. Very often the dollar values of a collection of items, such as the property records, are highly skewed, i.e., there is a relatively large number of small valued items, and a relatively small number of extremely large valued items. The distribution of an estimator based on a small sample size from such a population is usually skewed as well. Hence, it is not normal.

In the next section of this report we present the results of a simulation study that looks at the distribution of an estimator of the total in-place cost. The results suggest that the distribution of the estimator is not normal, and that the Satterthwaite approximation does not provide enough of an adjustment for the degrees of freedom. We present an alternative derived from the simulation results.

More advanced techniques such as balance repeated replication, or the jackknife can be used to determine error bounds when more complex situations arise.[9] We will not go into these methods here, since we believe our point about the increase in the size of the margin of error has been made.

[9] See Cochran, chapter 11, sections 18 - 20. See also, Wolter, K. M. (1985). *Introduction to Variance Estimation*. Springer-Verlag, New York.

Finally, once a correct approach is found for calculating error bounds, we would argue that a one-sided lower confidence bound should be used as the value assessed to be in error, e.g., the total in-place cost of non-locatable line-items, or the proportion of non-compliant records. This is because only values smaller than the lower bound are, statistically speaking, significantly different from values above the lower bound.[10] As noted in the main summary, the IRS uses such a rule for its audit findings.[11] Also, if one is going to take a conservative approach, the confidence level for this bound should be set at 99 percent. This practice attempts to take into account the uncertainty caused by various unquantifiable errors introduced into both the sampling and audit processes. In other words, as far as the FCC estimates of dollar values are concerned, use of the lower bound of the proper confidence interval would be the prudent approach.

Sources of Bias that Affect the Estimates

Several forms of bias are present in the estimates supplied in the draft report. These include:

- the use of a statistically biased estimator,
- bias caused by substituting CLLI and line-items for undesirable ones that turned up in the sample, and
- biases induced by a lack of control procedures in the audit process.

The effect of each of these biases varies in its degree of severity. The total effect is significant, and brings up legitimate concerns for the accuracy of the audit results. We address each in turn below.

Estimator Bias

The estimator used by the FCC is statistically biased. It can be a useful estimator in many situations, and has a smaller mean squared error than the standard unbiased estimator. The formula for the estimator of a total population value is given by

[10] This concept comes from the statistical theory of hypothesis testing. Suppose, for example, that a standard of "materiality" was set for the audit as should have been done. In particular, suppose it can be argued that there would be no cause for concern if the true in-place cost associated with non-locatable line-items is 100 million dollars or less (a very small fraction of the hardwire total investment). Then one would test the null hypothesis that the true value is less than or equal to 100 million dollars versus the alternative that there is more than 100 million dollars associated with non-locatable line-items. The deciding factor of this test is whether or not a one-sided lower confidence bound is below or above whatever materiality standard is set. If it is below, then one could not statistically conclude that the true value is materially different from the value carried on the CPR.

[11] See footnote 5 or the summary report for which this appendix is an attachment.

$$\hat{Y}_{R}=\sum_{h=1}^{L}\frac{M_h}{M'_h}\sum_{i=1}^{n_h}\frac{M_{hi}}{36}\sum_{j=1}^{36}y_{hij}=\sum_{h=1}^{L}\frac{M_h}{M'_h}\sum_{i=1}^{n_h}M_{hi}\bar{y}_{hi}=\sum_{h=1}^{L}\hat{Y}_{Rh}\text{, where}$$

$$\hat{Y}_{Rh}=\frac{M_h}{M'_h}\sum_{i=1}^{n_h}M_{hi}\bar{y}_{hi}\text{, and}$$

$$\bar{y}_{hi}=\frac{1}{36}\sum_{j=1}^{36}y_{hij}\,.$$

If y_{hij} is the in-place cost of an audited line-item that was not located and zero otherwise, then $\hat{Y}_R$ is an estimator for the total in-place cost of non-locatable line-items. On the other hand, if y_{hij} is one or zero depending on whether or not an audited line-item is or is not compliant, then $\hat{Y}_R$ is an estimator for the total number of compliant line-items in the population. If this is divided by M_0 then we have an estimator for the proportion of compliant line-items in the population.

In order to judge the precision of an estimator, statisticians usually look at the variance of the estimator, or its square root – the standard error of the estimator, For a biased estimate, the variance does not capture the precision of the estimator with respect to the true value of the population that is being estimated. The more appropriate measure is the mean squared error of the estimator, and its square root – referred to as the root mean squared error.

An approximate sample estimate for the mean-squared error for this estimator is given by

$$v(\hat{Y}_R)=\sum_{h=1}^{L}\left(\frac{N_h^2(1-f_{1h})}{n_h}\cdot\frac{\sum_{i=1}^{n_h}M_{hi}^2\left(\bar{y}_{hi}-\bar{\hat{Y}}_{Rh}\right)^2}{n_h-1}+\frac{N_h}{n_h}\sum_{i=1}^{n_h}\frac{M_{hi}^2(1-f_{2hi})s_{2hi}^2}{36}\right)\text{, where}$$

$$\bar{\hat{Y}}_{Rh}=\frac{\hat{Y}_{Rh}}{M_h},$$

$$f_{1h}=\frac{n_h}{N_h},\quad f_{2hi}=\frac{36}{M_{hi}}\text{ and}$$

$$s_{2hi}^2=\frac{1}{35}\sum_{j=1}^{36}\left(y_{hij}-\bar{y}_{hi}\right)^2.$$

This approximation depends on how well the ratio $\frac{M_h}{M'_h}$ approximates the ratio $\frac{N_h}{n_h}$ for each h = 1,…,L. This will depend on how much the M_{hi} vary within each stratum, and by how large n_h is for each stratum. If these ratio approximations are not good, then this formula produces a significantly biased estimate of the mean squared error. We can compare these numbers across all strata by looking at the total squared difference between the two. The square root of this total is the Euclidean distance between the two vectors, so this gives us a way to measure the closeness of the ratios across strata. The table below gives results.

Stratum h	Bell Atlantic North $\frac{M_h}{M'_h}$	$\frac{N_h}{n_h}$	Squared Error	Bell Atlantic South $\frac{M_h}{M'_h}$	$\frac{N_h}{n_h}$	Squared Error
1	4.119	4.000	0.014	5.720	5.000	0.518
2	10.348	9.833	0.265	15.752	15.250	0.252
3	15.695	16.000	0.093	14.014	14.000	0.000
4	15.151	15.500	0.122	15.293	15.000	0.086
5	12.373	13.000	0.393	21.201	21.000	0.040
6	20.621	21.000	0.144	25.734	26.000	0.071
7	32.194	30.000	4.814	39.549	43.000	11.909
8	42.453	43.000	0.299	38.109	39.667	2.427
9	32.064	30.333	2.996	51.338	55.500	17.322
10	107.320	112.500	26.832	65.228	69.000	14.228
11	127.228	113.000	202.436	141.867	130.333	133.033
12	N/A	N/A	N/A	281.014	233.667	2241.738
Total			238.408			2421.624

For each company, there are one or two stratum that contribute to most of the total squared difference. This gives us an indication that the approximation for the mean squared error of the biased estimator, $v(\hat{Y}_R)$, may not be very good.

To further evaluate the statistical bias in each of $\hat{Y}_R$ and $v(\hat{Y}_R)$, we conducted a simulation experiment that estimated the total in-place cost of the each population under study using. This was done as follows:

1. Define a frame of CLLI for which the total number of line-items, and the total in-place cost is know. Then divide the CLLI into 11 or 12 stratum depending on which Bell Atlantic company is being studied. For a summary of the frames we used, and how they compare to the frames used by the FCC for the audits, see the next section below.
2. Randomly select n_h out of the N_h CLLIs within each stratum, and record c_{hi}, the total in-place cost for selected CLLI i in stratum h.
3. Estimate the total in-place cost using

$$\hat{e}_R = \sum_{h=1}^{L} \frac{M_h}{M'_h} \sum_{i=1}^{n_h} C_{hi}$$

where C_{hi} is the total in-place cost for CLLI i within stratum h.

4. Estimate the mean squared error of $\hat{C}_{\mathrm{R}}$ using

$$v(\hat{C}_{\mathrm{R}}) = \sum_{h=1}^{\mathrm{L}} \left(\frac{\mathrm{N}_h^2 (1 - f_{1h})}{\mathrm{n}_h} \cdot \frac{\sum_{i=1}^{\mathrm{n}_h} \mathrm{M}_{hi}^2 \left(\bar{\mathrm{c}}_{hi} - \bar{\bar{\mathrm{C}}}_{\mathrm{R}h} \right)^2}{\mathrm{n}_h - 1} \right), \text{ where}$$

$$\bar{\mathrm{c}}_{hi} = \frac{1}{\mathrm{M}_{hi}} \mathrm{C}_{hi}$$

$$\bar{\bar{\mathrm{C}}}_{\mathrm{R}h} = \frac{1}{\mathrm{M}'_h} \sum_{i=1}^{\mathrm{n}_h} \mathrm{C}_{hi} \text{ , and}$$

$$f_{1h} = \frac{\mathrm{n}_h}{\mathrm{N}_h}.$$

5. Repeat steps 1 through 4 a large number of times. In our case we did 5,000 runs.

While this simulation does not perform an evaluation using $\hat{Y}_{\mathrm{R}}$ to estimate values the audit is interested in, it does provide information about how well $\hat{Y}_{\mathrm{R}}$ and $v(\hat{Y}_{\mathrm{R}})$ perform in estimating the in-place cost associated with non-locatable line-items. This is because the simulation looks at estimates of a similar quantity, total in-place cost.

For evaluating the bias of $\hat{C}_{\mathrm{R}}$, we compared the average of the 5,000 realizations with the know value of the total in-place cost.

Item	Bell Atlantic North	Bell Atlantic South
Total Hardwire In-place Cost	\$5,715,386,027	\$6,435,487,414
Average value of $\hat{C}_{\mathrm{R}}$	5,703,490,468	6,427,576,434
Standard Error of the Average of $\hat{C}_{\mathrm{R}}$	5,026,648.14	5,985,124.03
Estimated Bias	-11,895,559	-7,910,980
Bias as a Percentage of the Total	-0.21%	-0.12%

These results indicate that the bias of the estimator $\hat{C}_{\mathrm{R}}$ is not that bad. The mean value of the estimator is approximately one to two tenths of a percent of below the actual total hardwire investment.

We now evaluate the bias in the approximation of the mean squared error of $\hat{C}_{\mathrm{R}}$. To do this we first estimate the mean squared error in the following way:

1. For each of the 5,000 realizations of $\hat{C}_{\mathrm{R}}$, subtract the true total in-place cost from the estimate.
2. Square each of these errors.

3. Find the average of the 5,000 squared errors.

We then compared the result of this calculation with the average of the 5,000 values of $v(\hat{e}_R)$.

Item	**Bell Atlantic North** Squared $ $\times 10^{15}$	**Bell Atlantic South** Squared $ $\times 10^{15}$
Simulation Estimate of the Mean Squared Error	126.452	179.135
Standard Error of the Simulation Estimate	3.511	4.705
Average value of $v(\hat{e}_R)$	121.472	175.225
Standard Error of the Average of $v(\hat{e}_R)$	2.027	3.213
Estimated Bias	-4.980	-3.910
Standard Error of the Estimated Bias	2.802	3.754
Bias as a Percentage of the Total	-3.94%	-2.18%

These results are not necessarily conclusive, but it appears that $v(\hat{e}_R)$ underestimates the mean squared error of $\hat{e}_R$ enough for there to be some concern. If one felt that this is significant, then an adjustment should be made to $v(\hat{e}_R)$. We suggest inflating the estimated mean squared error for the North by 4 percent and the South by 2 percent.

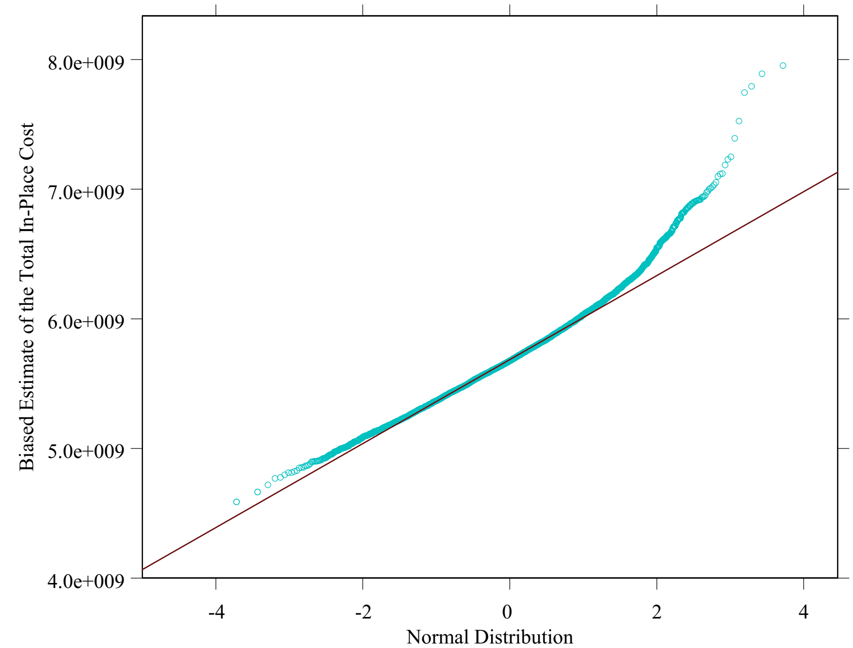
The simulation results can also give us an indication of how to proceed with determining a one-sided lower confidence bound. We can use the results to answer the following questions.

1. Is the underlying distribution of $\hat{e}_R$ normally distributed?
2. Can Student's t distribution be used to find the multiplying factor for determining a lower confidence bound?
3. Is the Satterthwaite approximation for the effective degrees of freedom good?

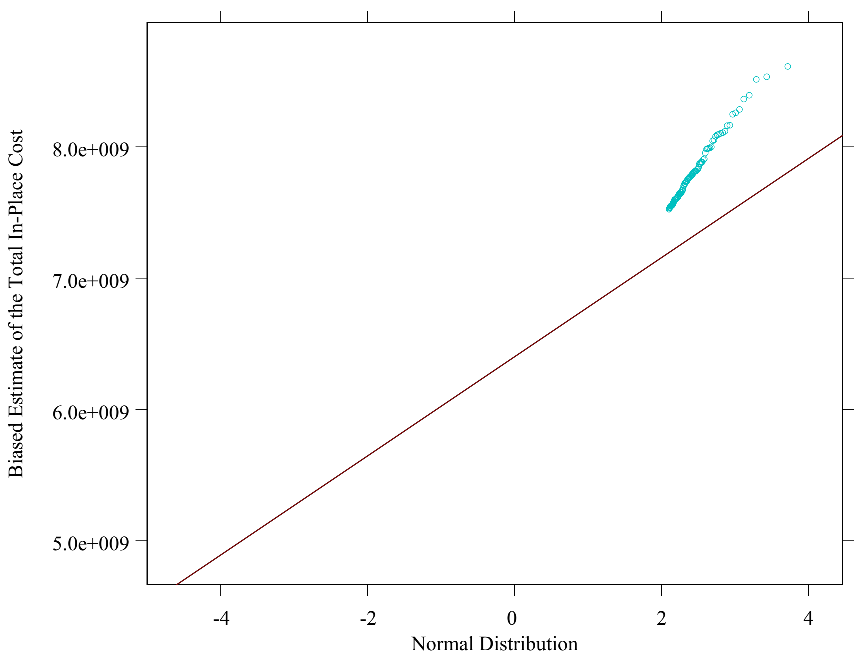
To answer the first question, we looked at a normal q-q plot of the 5,000 realizations of $\hat{e}_R$. This plot provides a powerful, visual comparison of the estimated quantiles of $\hat{e}_R$ with the same quantiles of a standard normal distribution. If $\hat{e}_R$ has a normal distribution, then the resulting plot should look like a straight line. We follow Cleveland's[12] method of presentation where a reference line passing through upper and lower quartiles is superposed on the graph.

[12] Cleveland, W. S. (1993) *Visualizing Data*. Hobart Press, Summit, New Jersey.

Normal Q-Q Plot for Estimated Values of $\hat{\theta}_R$
Bell Atlantic North



Normal Q-Q Plot for Estimated Values of $\hat{\theta}_R$
Bell Atlantic South



These plot conclusively show that is $\bar{C}_R$ <u>not</u> normally distributed. As we conjectured, values are highly skewed. Thus, once should not used normal distribution theory to find confidence bounds.

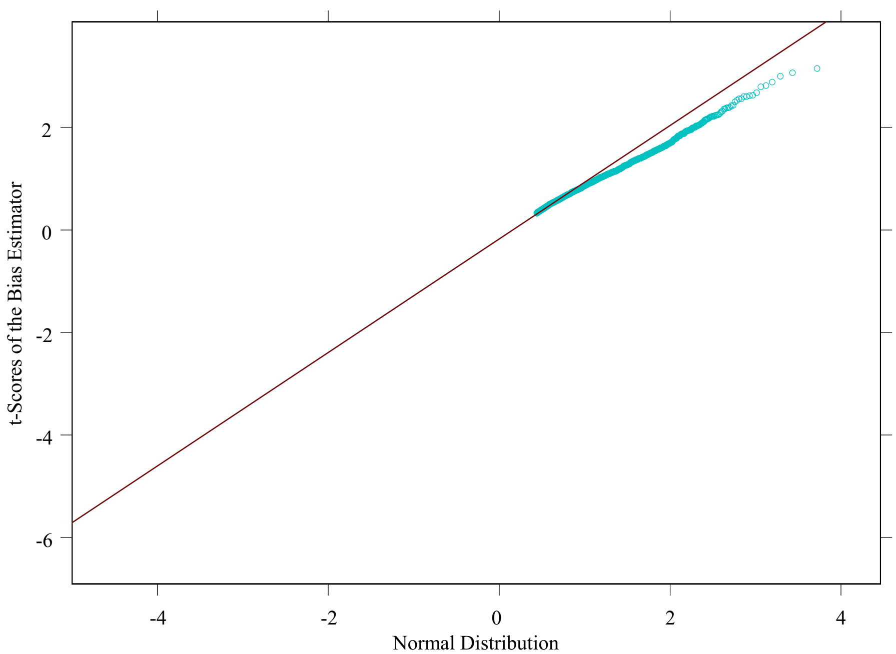
Given that, is still possible to find a multiplier for the root mean squared error that will appropriate margins of error?

To determine this we compute a t-score of $\bar{C}_R$ from each realization in the simulation, and look at the distribution of these t-scores. To compute t-scores we do the following.

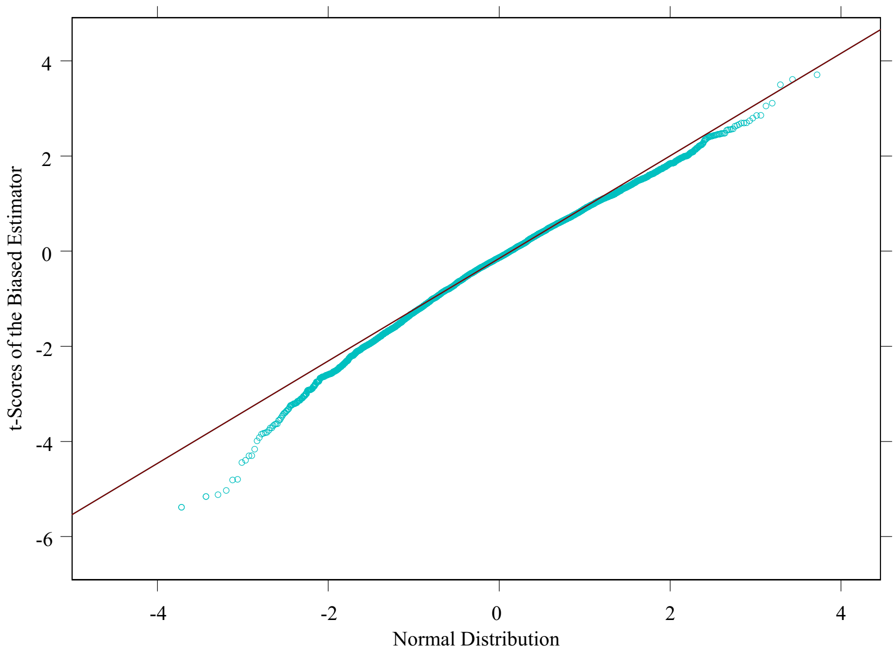
1. Calculate the error in each estimate, $\bar{C}_R$ - C, where C is the known total in-place cost.
2. Divide the error by $\sqrt{v(\bar{C}_R)}$ or $\sqrt{\varphi \cdot v(\bar{C}_R)}$ where φ is the inflation factor for the MSE discussed above.

We examine the distribution of these t-scores first by comparing it with a normal distribution via a normal q-q plot. Only the t-score using the uninflated mean squared error estimate are show. The plots of t-scores using the inflated mean squared error estimate will have the same general shape.

Normal Q-Q Plot for t-Scores of $\bar{C}_R$
Bell Atlantic North



Normal Q-Q Plot for t-Scores of ∂_R
Bell Atlantic South



These plots tell us that the tail of the distribution for negative t-scores is much heavier than that of a normal distribution – much like Student's t distribution. However, the other tail is similar to a normal for the South, but much thinner than a normal for the North.

To find multipliers for the root mean squared error in order to obtain one-sided lower confidence bounds, we can use the 1 percent or 5 percent quantiles of the t-score distribution. These are presented below.

	Bell Atlantic North		**Bell Atlantic South**	
	1%	5%	1%	5%
t-score	-3.213	-2.244	-3.107	-2.093
t-score (inflated MSE)	-2.936	-2.051	-3.074	-2.071

We can also use the simulation results to see how good the Satterthwaite approximation is for finding an effective degrees of freedom. Recall, that the approximation is only known to be good when the underlying distribution of the estimator is normally distributed. This is not the case here. But since the left tail of the t-scores behaves somewhat like a t distribution, the Satterthwaite effective degrees of freedom may be reasonable for use in calculating lower confidence bounds.

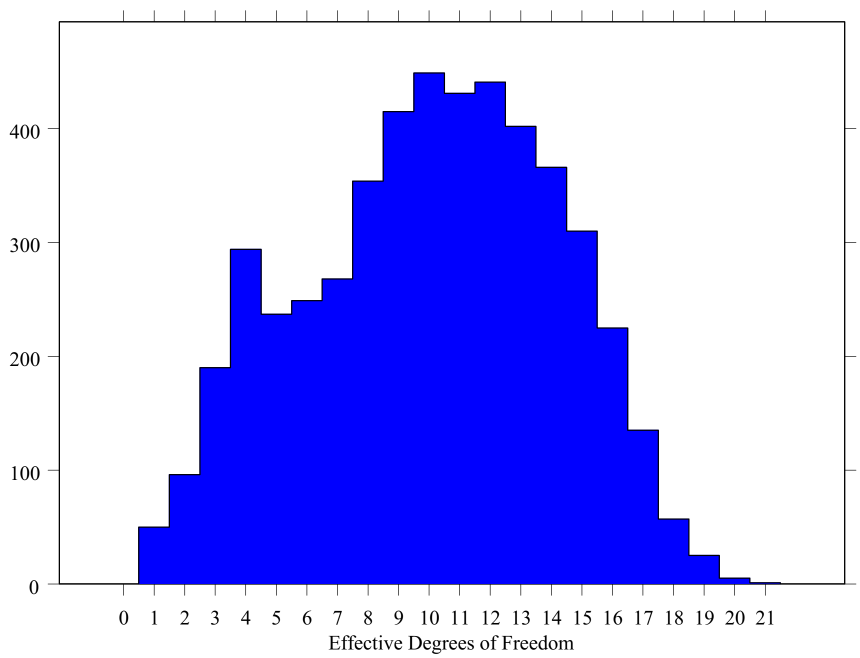
We proceed by first identifying the degrees of freedom for a Student's t distribution that fits the lower tail of the uninflated t-scores. We do this by finding a least squares fit between the quantiles of the negative t-score and a t distribution.

- For Bell Atlantic North, a t distribution with 5.397 degrees of freedom provides a least squares fit.
- For Bell Atlantic South, a t distribution with 12.572 degrees of freedom provides a least squares fit.

Next, we calculate the effective degrees of freedom for each of the 5,000 realizations in the simulation. We can then compare the distribution of the effective degrees of freedom with the least squares fit to see if the effective degrees of freedom provides similar results.

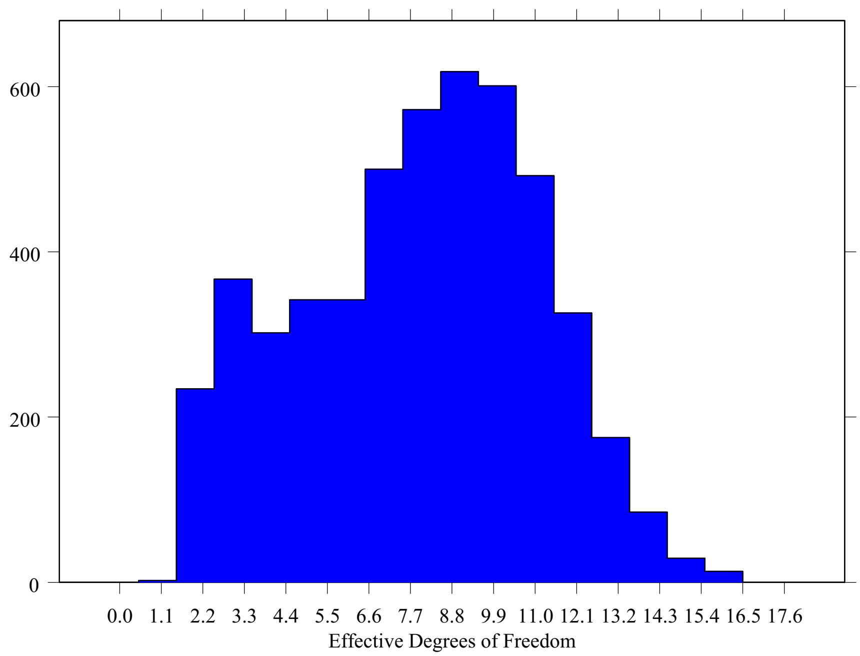
For Bell Atlantic North, the distribution of the effective degrees of freedom is quite wide. It has a mean of 10.045, and a standard deviation of 4.102. A histogram of the distribution is shown below.

Histogram of 5,000 Realizations of the Effective Degrees of Freedom
Bell Atlantic North



For Bell Atlantic South, the distribution of the effective degrees of freedom is also wide. It has a mean of 7.928, and a standard deviation of 3.138. A histogram of the distribution is shown below.

Histogram of 5,000 Realizations of the Effective Degrees of Freedom
Bell Atlantic South



In neither case does the effective degrees of freedom calculation appear to capture the t distribution that best fits the left tail of the corresponding t-score distribution. This should not be surprising, since we do not have an underlying normal distribution to start with. Note, that the least squares fits to the t-score distributions are not necessarily a proper fit of the left tails, but it does give us a way to evaluate the effective degrees of freedom.

In light of the simulation results, we will compute lower bounds for biased estimates of dollar values using the appropriate quantile of the t-score distribution given above. We do not know if this analysis is compatible with proportion estimation. When estimating proportions we will use the Satterthwaite effective degree of freedom to find the confidence bound multiplier.

<u>Substitution Biases</u>

Aside from statistical calculation issues, there are other sources of bias. We are concerned about statements in the reports that refer to substitutions of sampled items. First, both reports state on page 7 of Appendix B that

> "In some instances, the location initially selected was impractical to audit, …. In such cases, another location was randomly selected from that stratum."

If the FCC does not want to audit certain locations, their conclusions should be narrowed accordingly – in fact, just to the records in locations that the FCC was willing to audit.

Additionally, the report talks about substituting for "hard-to-get-to" line items by using the preceding line-item in the population list. This changes the probability of selection for certain items, and if this is not accounted for the result is bias in the estimate.

<u>Biases Induced by a Lack of Control Procedures</u>

It appears that there was a lack of control procedures throughout the audit process. For example, the audit staff did not use the same team of auditors to inspect each location. When examining the proportion of items found by different audit teams, there are noticeable differences in the scoring of line-items. The audit staff tried to correct this control problem by making extensive "back-at-the-office" changes in the scores. It is unclear whether they succeeded in addressing the original team variability in approach since no locations were revisited to verify that the back office scoring correctly represents the true state of the property records.

The FCC has twice changed the initial on-site audit scores. These changes were not based on additional information or revisits. They were made back in the office after supervisory review to compensate for inconsistencies in coding.

In the first set of code changes, over 15 percent of the South's initial scores were changed based on solely on reviewer's notes. The current set of FCC codes had an additional 194 changes (based on combined Bell Atlantic North and South data sets). This is an additional 8 percent change on top of the original revisions. Most of the re-coded items were scored as a "4" on the last version of FCC coding. Roughly half of these were re-scored as "1" and the other half as "3". The increase in score "3" items has serious consequences on the estimate of the misrepresented costs.

With this many revisions, there is cause to suspect the accuracy of the entire scoring process. The FCC may argue that the revisions improve consistency of coding and correct errors. The point is, if there were this many errors or inconsistencies in the first place, how can we be assured all errors and inconsistencies were identified and corrected? Also, were some of the last revisions a result of a policy change in the categorization of certain items previously scored as a "4"? If so, how can we be certain the coding is consistent when the coding criteria has changed long after the onsite FCC visits – nearly a year after in fact?

The error introduced by incorrect scores is not accounted for in statistical estimates, variance equations, or confidence interval calculations. This is *non-sampling* error because it comes from a source other than the random selection of a sample. It is difficult to quantify non-sampling error. However, had the previous set of FCC codes been used (using the unbiased estimator discussed below), the estimate of misrepresented costs would have been $231 million with a standard error of $55 million. Using the current scores, the estimate is $436 million with a standard error of $94 million. Clearly, the re-scoring had a substantial impact.

Calculations

Step 4 of the sample design process was carried out on both the North and South CPR databases that were given to us by Bell Atlantic. We were unable to produce a sampling frame that produced a summary table exactly the same as those given in Appendix B page 7 of each of the draft reports.

<u>Calculation Comparisons</u>

We went ahead with the calculation of estimates and variances based on sampling frames that exhibit many of the same summary characteristics as those of the summaries given in the draft reports. We based these "best match" frames on the summary spreadsheets that Bell Atlantic gave to the FCC as part of Data Request 2. This summarized in Table as taken from the each of the draft reports, along with the summary of the "Best Match" frame we used in our calculations (shown in Table 2).

Notice that the N_h and n_h match up fairly well. The record counts within strata do not necessarily match very well. This will affect our calculations, and we will not be able to judge whether or not the equations for the biased estimator and its mean squared error were truly the ones used by the FCC.

Tables 3 through 6 the results of calculations published in the draft reports with our attempt to verify the numbers.

Our calculations for the South's proportion estimate are fairly good. All of our other calculations do not match-up well with those published in the reports. We suspect this is mainly due to the fact that our record counts within CLLI do not match well with those that the FCC used.

<u>Updated Results</u>

We now present the results (in Tables 7 and 8) of calculating the estimates under four different scenarios:

1. using the biased estimator and its approximate mean squared error with the FCC's current scoring of audited line items (given in the draft report);
2. using the unbiased estimator and its variance with the FCC's current scoring of audited line items (given in the draft report);
3. using the biased estimator and its mean squared error with the Bell Atlantic's scoring of audited line items; and
4. using the unbiased estimator and its variance with the Bell Atlantic's scoring of audited line items.

The unbiased estimator is one that is based on weighting each value in the sample by the inverse of the probability of selection. As with the biased estimator, it can be used to estimate the proportion of compliant line-items, or the total in-place cost of non-locatable line-items. The formula for the estimator and its variance can be found in Cochran.

One-sided lower (upper) bounds are given for the in-place cost (proportion) estimates. If, for example, the one-sided lower bound the in-place cost of non-locatable line-items was less than 100 million dollars, then one could not statistically conclude, at the appropriate confidence level, that the true value is actually more than 100 million dollars. Similarly, if the one-sided upper confidence bound is greater than 0.90, then one could not statistically conclude, at the appropriate confidence level, that the true value is actually less than 0.90 (90 percent).

To calculate one-sided lower confidence bounds for all proportion estimates and the unbiased in-place cost estimates, Student's t distribution was used with the effective degrees of freedom calculated using the Satterthwaite approximation. For the biased estimates of in-place cost, the multiplying factors determined by the simulation results in the previous section were used.

Table 1 - FCC vs. "Best Match" (BM) Sampling Frame Summary
Bell Atlantic North

Stratum *h*	**Number of Records per Central Office Location in Stratum** High **FCC**	**BM**	Low **FCC**	**BM**	**Number of Central Office Locations in Stratum** N_h **FCC**	**BM**	**Number of Central Office Locations Selected for Audit** n_h **FCC**	**BM**	**Number of Records per Stratum** M_h **FCC**	**BM**
1	6801	6801	2033	2033	23	24	6	6	75513	74926
2	1985	1986	1008	994	63	59	6	6	85902	81489
3	987	987	809	809	31	32	2	2	27254	28063
4	796	796	701	701	29	31	2	2	21573	23075
5	696	696	601	601	28	26	2	2	18140	16889
6	599	599	500	500	54	63	2	3	29568	34458
7	499	499	400	400	65	60	3	2	29543	27107
8	498	398	301	301	82	86	2	2	28729	30184
9	296	299	200	200	91	91	2	3	22077	22188
10	199	199	100	100	452	450	4	4	60538	60743
11	99	99	79	79	228	226	3	2	20234	20102
Total					1146	1148	34	34	419071	419224

Table 2 - FCC vs. "Best Match" (BM) Sampling Frame Summary
Bell Atlantic South

Stratum *h*	**Number of Records per Central Office Location in Stratum** High **FCC**	**BM**	Low **FCC**	**BM**	**Number of Central Office Locations in Stratum** N_h **FCC**	**BM**	**Number of Central Office Locations Selected for Audit** n_h **FCC**	**BM**	**Number of Records per Stratum** M_h **FCC**	**BM**
1	5303	5303	2027	1974	20	20	4	4	59074	53508
2	1973	1973	1002	1002	59	61	4	4	77069	77435
3	995	1001	902	902	31	28	2	2	29564	26668
4	895	899	801	800	31	30	2	2	26171	25326
5	799	799	710	704	42	42	2	2	31326	31314
6	699	702	603	603	76	78	3	3	49102	50464
7	599	600	501	501	88	86	2	2	48124	47103
8	499	499	400	400	115	119	3	3	52516	53391
9	398	399	300	300	113	111	2	2	39470	38709
10	299	299	200	200	144	138	2	2	35624	34375
11	199	199	100	100	384	391	3	3	52358	53484
12	99	99	36	36	697	701	3	3	41142	41309
Total					1800	1805	32	32	541540	533086

Table 3 - Calculation Verification
Percentage of Compliant Records
Bell Atlantic North

Item	Estimate	Standard Error	Margin of Error[13]	Lower Confidence Bound[13]	Upper Confidence Bound[13]
FCC Published	79.48	2.00	3.92	75.56	83.40
Verification Results	79.50	2.36	4.62	74.88	84.12

Table 4 - Calculation Verification
Total In-Place Cost ($M) of Non-Locatable Line-Items
Bell Atlantic North

Item	Estimate	Standard Error	Margin of Error[13]	Lower Confidence Bound[13]	Upper Confidence Bound[13]
FCC Published	404.6	81.8	160.4	244.2	565.0
Verification Results	441.4	85.0	166.7	274.7	608.1

Table 5 - Calculation Verification
Percentage of Compliant Records
Bell Atlantic South

Item	Estimate	Standard Error	Margin of Error[13]	Lower Confidence Bound[13]	Upper Confidence Bound[13]
FCC Published	75.52	1.95	3.82	71.70	79.34
Verification Results	75.54	1.96	3.83	71.71	79.38

Table 6 - Calculation Verification
Total In-Place Cost ($M) of Non-Locatable Line-Items
Bell Atlantic South

Item	Estimate	Standard Error	Margin of Error[13]	Lower Confidence Bound[13]	Upper Confidence Bound[13]
FCC Published	888.9	185.8	364.3	524.6	1253.2
Verification Results	847.8	178.6	350.1	497.7	1197.8

[13] The margin of error was determined using the 97.5th percentile of a standard normal distribution. This is the same methodology used in the draft report. While we do not feel that this is the correct thing to do, it was done here in order to make a comparison. The resulting error bounds give a 95 percent confidence interval.

Table 7 - Percent of Compliant Records

Bell Atlantic Company	Scoring System	Estimator Type	Estimate	Standard Error	Effective Degrees of Freedom	One-Sided 95% Upper Bound	One-Sided 99% Upper Bound
North	FCC	Biased	79.50	2.36	13.91	83.67	85.74
		Unbiased	78.40	3.76	9.43	85.30	89.02
	Bell Atlantic	Biased	86.62	1.76	10.30	89.82	91.50
		Unbiased	85.45	3.64	8.40	92.23	96.01
South	FCC	Biased	75.54	1.96	12.42	79.03	80.79
		Unbiased	74.43	2.24	13.41	78.39	80.36
	Bell Atlantic	Biased	77.26	1.83	12.89	80.52	82.16
		Unbiased	76.03	2.17	13.18	79.87	81.78

Table 8 - Total In-Place Cost ($M) of Non-Locatable Line-Items

Bell Atlantic Company	Scoring System	Estimator Type	Estimate	Standard Error	Effective Degrees of Freedom	One-Sided 95% Lower Bound	One-Sided 99% Lower Bound
North	FCC	Biased	441.4	85.0	N/A	250.6	168.2
		Unbiased	436.2	94.6	7.77	257.1	152.7
	Bell Atlantic	Biased	71.0	23.2	N/A	18.8	-3.7
		Unbiased	69.8	24.7	10.36	25.0	1.6
South	FCC	Biased	847.8	178.6	N/A	473.9	292.8
		Unbiased	839.8	196.6	5.77	443.6	178.2
	Bell Atlantic	Biased	217.5	89.9	N/A	29.5	-61.7
		Unbiased	212.3	97.0	1.48	-400.3	-2875.0

Appendix B
Bayesian Methodology

In this Appendix, we provide some insight into the general Bayesian approach for analyzing data sampled from a finite population. We also briefly discuss the issue of unbiased estimators. Finally, we give an example for which the distribution of the total cost of the hard-wired COE not found, under a Bayesian model approach, is a sum of independent random variables, one of which is Cauchy distributed. This result reiterates our earlier (frequentist-based) concern over the understated margin of error.

Bayesian Approach

Before setting up the Bayesian model for the total cost, we take a look at the general Bayesian approach for sampling. Let $y = (y_1, \Lambda\ , y_N)$ denote the data of the finite population of interest. The superpopulation refers to a hypothetical population from which y can be thought to have been sampled. The distribution of y in the superpopulation has parameters $\boldsymbol{\theta}$. Suppose a simple random sampling technique was employed to sample n observations from the finite population. Denote the observed sample of y as y_{obs}, and the unobserved part of y as y_{unobs}. Suppose we are interested in estimating the finite population average, $\bar{y}$, which can be calculated as

$$\bar{y} = \frac{n}{N}\bar{y}_{obs} + \frac{N-n}{N}\bar{y}_{unobs}. \qquad (1)$$

The typical Bayesian approach proceeds as follows.

1. Pick a likelihood for the observed data

$$p(y_{obs}|\boldsymbol{\theta}) = \prod_{i=1}^{n} p(y_{obs_i}|\boldsymbol{\theta}).$$

Put a prior distribution $p(\theta)$ on the parameters θ. Calculate the posterior distribution of θ

$$p(\boldsymbol{\theta}|y_{obs}) \propto p(\boldsymbol{\theta})\ p(y_{obs}|\boldsymbol{\theta}).$$

2. Simulate $\theta^l, l = 1, \Lambda\ , L$, from $p(\theta|y_{obs})$. For each θ^l draw a vector y_{unobs} from

$$p(y_{unobs}|\boldsymbol{\theta}^l) = \prod_{j=1}^{N-n} p(y_{unobs_j}|\boldsymbol{\theta}^l),$$

and average over y_{unobsj} to obtain a draw of $\bar{y}_{unobs}$. Using equation (1), we produce a draw of $\bar{y}$. Averaging over the total L draws of $\bar{y}$, we obtain an estimate of the finite population average $\bar{y}$

Another way of simulating $\bar{y}$ is directly from its posterior $p(\bar{y}|y_{obs})$. This can only be done in special cases when the choice of likelihood and prior result in a mathematically tractable posterior distribution function. In particular, to illustrate ideas, we could choose a normal likelihood with standard noninformative prior, then we have

$$\bar{y}|y_{obs} \sim t_{n-1}(\bar{y}_{obs}, (\frac{1}{n} - \frac{1}{N})s^2_{obs})\ [1]. \qquad (2)$$

---

[1] See Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin, *Bayesian Data Analysis,* Chapman and Hall, 1998, page 203.

Unbiased Estimates

Under frequentist theory, the concept of unbiasedness is that, over repeated sampling, the average of a parameter estimate should be equal to its true value. This is usually considered an intuitively appealing concept, and much theory has been developed concerning minimum variance unbiased estimators.

In the Bayesian world, one wants to determine the (posterior) distribution of a parameter; the concept of an unbiased point estimate is unimportant. The principle of unbiasedness is reasonable in the limit of large samples, but otherwise is potentially misleading.[2]

In Bayesian sampling theory terms, minimizing bias will often lead to counterproductive increases in variance. Thus, the FCC staff's statement that since the Bayesian method is design free, the estimator is unbiased seems to us to be irrelevant to the Bayesian analysis that was performed.

Example

To estimate the total cost of COE not found, we assume that for each selected central office, all line-items in the office are sampled and the corresponding cost of COE not found are observed. The sample design is simplified to a one stage stratified design. For each strata, let $x_j = (x_{1j}, \Lambda, x_{n_j j})$, where $x_{ij} = 1$ when central office i is in stratum j and 0 otherwise. The vector x_j is fully observed as long as we know the total number of central offices N_j in the stratum and the values of x_{ij} for the central office in the sample. The variables x_j incorporate the stratification information, and the Bayesian model is conditioned on x_j. For each strata, let $y_j = (y_{1j}, \Lambda, y_{N_j j})$ be the set of population values of the cost of COE not found. We model the distribution of y_j within stratum j in terms of parameters θ_j. Under the assumption that the eleven vectors of parameters $(\boldsymbol{\theta}_1, \Lambda, \boldsymbol{\theta}_{11})$ have independent and identically distributed priors, we can obtain posterior inferences separately for the parameters in each stratum. Suppose further that for each strata we choose the likelihoods, $p(y_{ij} | x_{ij}, \boldsymbol{\theta}_j)$, to be normal and the priors, $p(\theta_j)$, to be the standard noninformative ones, then using result (2), we have

$$\bar{y}_j | y_{obs\,j} \sim t_{n_j - 1}(\bar{y}_{obs_j}, (\frac{1}{n_j} - \frac{1}{N_j}) s^2_{obs_j}), \tag{3}$$

where $\bar{y}_j$ is the average cost in stratum j. According to the FCC report, the total cost is calculated using

$$C = \sum_{j=1}^{11} N_j \bar{y}_j .$$ [3]

Applying result (3), we have

$$N_j \bar{y}_j | y_{obs\,j} \sim t_{n_j - 1}(N_j \bar{y}_{obs_j}, N_j^2 (\frac{1}{n_j} - \frac{1}{N_j}) s^2_{obs_j}). \tag{4}$$

---

[2] Ibid. Page 108.

[3] The audit sample design for Bell Atlantic North had eleven strata. The design for Bell Atlantic South consisted of twelve strata.

From result (4), we see that the total cost of COE not found, C, is just the sum of eleven independent random variables. In the FCC audit sample of Bell Atlantic North, six out of the eleven strata have two central offices selected. In this situation, the degrees of freedom is 1. A t distribution with 1 degree of freedom is also known as a Cauchy distribution, and the sum of six independent Cuachy random variables is also Cauchy distributed.

The Cauchy distribution is heavy tailed; the moments of the distribution do not exist -- including the mean and variance. Since C is the sum of 5 independent non-central t random variables and one independent Cauchy random variable, C will not have a mean nor a variance.

If the FCC staff used a Bayesian approach similar to this, and did not realize that they were dealing with a Cauchy distribution, then the credibility intervals may not have been properly calculated. Thus, as was the case in the staff's frequentist analysis, the margin of error in the FCC report is understated.